

Experience Rating with Poisson Mixtures

G. O. Brown[†]

Statistical Laboratory, Centre for Mathematical Sciences, Cambridge CB3 0WB, UK

S. P. Brooks

Statistical Laboratory, Centre for Mathematical Sciences, Cambridge CB3 0WB, UK

W. S. Buckley

CBA, Florida International University, Miami, FL 33199, USA

Summary. We present a mixture Poisson model for claims counts in which the number of components in the mixture are estimated by reversible jump MCMC methods.

Keywords: Reversible Jump MCMC, Poisson Mixture Modelling

1. Introduction

In this paper we consider a mixed Poisson model for count data arising in Group Life insurance. We present a Bayesian formulation to determine the number of groups in an insurance portfolio consisting of claim numbers or deaths. We take a non-parametric Bayesian approach to modelling this mixture distribution using a Dirichlet process prior and use reversible jump Markov chain Monte Carlo to estimate the number of components in the mixture. The physical interpretation of the model in this case is that the heterogeneity is assumed to be drawn from one of a finite number of possible groups, in proportion which will be estimated.

2. A Credibility Model for Heterogeneity

The data arise from 1125 groups insured through the whole or parts of the period 1982–1985 by a major Norwegian insurance company. There are $n = 72$ classes distinguished by occupation category. The i^{th} class has risk exposure E_i , and observed number of deaths, D_i . The data are also analysed in Haastrup (2000) and Norberg (1989). Let D_1, \dots, D_n denote the number of observed deaths in each insured group. Associated with each group is the exposure, denote E_1, \dots, E_n , respectively, which is a measure of the propensity of that group to produce claims/deaths. Let D^n denote the collection of all deaths for each group, where

$$D^n = \{D_1, \dots, D_n\}.$$

Similarly, let E^n denote the collection of all exposures for the group

$$E^n = \{E_1, \dots, E_n\}.$$

[†]Email:gob20@statslab.cam.ac.uk

Figure 1 shows a plot of the claim number for each group while Figure 2 shows the claim numbers normalized by their corresponding exposures.

The heterogeneity model is used to model differences in each of the n groups. For each group, the exposures are recorded and the resulting number of deaths or claims are then recorded for each group. At the first level, the number of claims for each group is assumed to follow a Poisson distribution with parameter $\lambda_i E_i$, for $i = 1, \dots, n$. Thus

$$D_i \sim \text{Poisson}(\lambda_i E_i), \quad i = 1, \dots, n.$$

We take a fully Bayesian approach and assume that the λ_i are IID and follow a Gamma distribution with parameters α and β ; that is,

$$\lambda_i \sim \text{Gamma}(\alpha, \beta),$$

where α and β are also assumed to be unknown.

The advantage of using such mixed distributions is that it allows for extra variation in the number of occurrences since

$$\mathbb{E}(D) = \mathbb{E}(\mathbb{E}(D|\lambda)) = \mathbb{E}(E\lambda) = E\alpha/\beta$$

and

$$\begin{aligned} \text{Var}(D) &= \mathbb{E}(\text{Var}(D|\lambda)) + \text{Var}(\mathbb{E}(D|\lambda)) \\ &= E\alpha/\beta + E^2\alpha/\beta^2 > \mathbb{E}(D). \end{aligned}$$

3. Extending the Basic Model

In an analysis based on the model described in Section 2, Haastrup (2000) assumes that each group has its own unique heterogeneity parameter λ , drawn from some distribution. Haastrup (2000) assumes that each class i has a unique heterogeneity parameter, denoted λ_i , and that the number of deaths D_i in this class follows a Poisson distribution with mean $\lambda_i E_i$. The classes are assumed to be mutually independent, given the heterogeneity parameters $\lambda_1, \lambda_2, \dots, \lambda_n$. Furthermore, he assumes that this distribution is identical for each group. In practice, large values of E_i will account for large values of D_i , which will lead to similar values of λ_i for each i .

In our analysis, we propose a mixture model formulation. We assume that D_i , given λ_j , has a Poisson distribution with mean $\lambda_j E_i$. We take a non-parametric Bayesian approach to modelling this mixture distribution using a Dirichlet process prior, and use reversible jump Markov chain Monte Carlo to estimate the number of components in the mixture. In this case, the physical interpretation of the model is that the heterogeneity is assumed to be drawn from one of k possible groups, in proportions w_1, \dots, w_k .

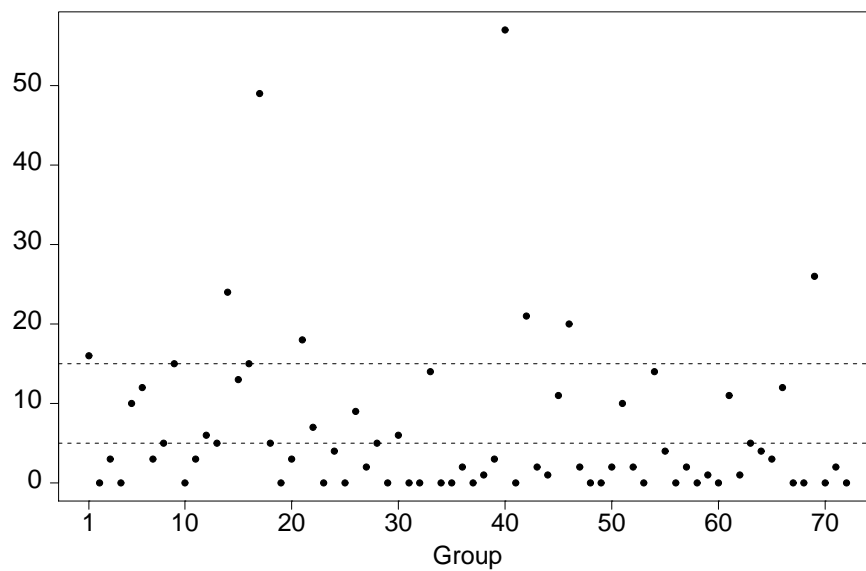


Fig. 1. Plot of the number of observed claims for each Group.

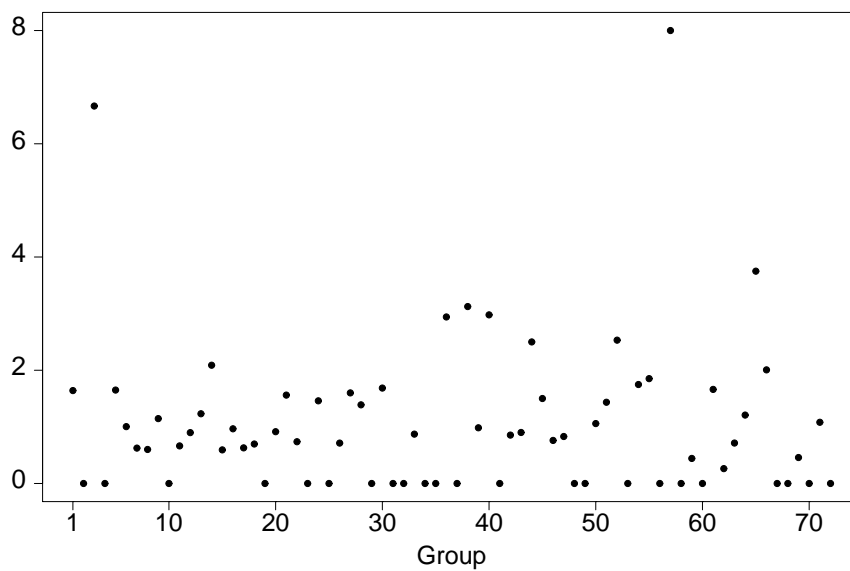


Fig. 2. Plot of the number of observed claims per unit exposure.

4. Mixture Formulation

The method we describe is essentially a classification problem where we assume that each observed D_i comes from any one of k components, where each component has a Poisson distribution. Thus

$$D_i | \lambda_j \sim \text{Poisson}(E_i \lambda_j) \quad j = 1, \dots, k; i = 1, \dots, n.$$

More general forms of the mixture Poisson model with covariates are discussed in Green and Richardson (2002). Mixture models for grouped claim numbers are considered by Tremblay (1992) and Walhin and Paris (1999, 2000). Dellaportas et al. (1997) considers count data in Finance using split/merge moves, while Viallefont et al. (2002) provides a more general discussion of mixtures of Poisson distributions, using both split/merge moves and birth/death moves. Other methods for determining the number of components in a mixture are discussed by McLachlan and Peel (2000), Phillips and Smith (1996), Carlin and Chib (1995), and Stephens (2000) who use Markov chains to model jointly the number of components and component values. The advantage of the Bayesian formulation is that we can place posterior probabilities on the order of the model.

4.1. The Likelihood Function

Throughout our discussion, n will denote the number of data points and k will denote the number of components in the mixture formulation. For a finite mixture model, the observed likelihood function is

$$L(D^n | \lambda, w, E^n) = \prod_{i=1}^n \sum_{j=1}^k w_j f_j(D_i | \lambda_j, E_i), \quad (1)$$

where the weights are non-negative and $\sum_{j=1}^k w_j = 1$. Even for moderate values of n and k , this takes a long time to evaluate since there are k^n terms when the inner sums are expanded (Casella et al., 2000). Another form of the likelihood function will be derived shortly. Classical estimation procedures for mixture models are described by Titterton et al. (1990) and McLachlan and Peel (2000).

Let z_i be a categorical random variable taking values in $\{1, \dots, k\}$ with probabilities w_1, \dots, w_k , respectively, so that

$$p(z_i = j | w) = w_j.$$

Suppose that the conditional distribution of D_i , given $z_i = j$, is $\text{Poisson}(\lambda_j)$, $j = 1, \dots, k$. Let $f_j(\cdot)$ denote a Poisson density with parameter λ_j . Then the unconditional density of D_i is given by $f(D_i)$, where

$$f(D_i) = \sum_{j=1}^k w_j f_j(D_i | \lambda, E^n),$$

since

$$\begin{aligned} f(D_i) &= \sum_{j=1}^k f_j(D_i|z_i = j, \lambda, E^n) p(z_i = j) \\ &= \sum_{j=1}^k w_j f_j(D_i|\lambda_j E_i). \end{aligned} \quad (2)$$

With every pair (D_i, E_i) , we associate a latent variable z_i , which is an indicator variable that indicates which component of the mixture is associated with (D_i, E_i) . We have, $z_i = j$ if the i^{th} data point, (D_i, E_i) , comes from the j^{th} component of the mixture. Thus, for each i , we have

$$z_i|w \sim \mathcal{M}(1; w_1, \dots, w_k)$$

and

$$D_i|z_i \sim \mathcal{P}(\lambda_{z_i} E_i).$$

By incorporating the indicator variables z , the complete data likelihood is then

$$\begin{aligned} L(D^n|z, \lambda, E^n) &= \prod_{i=1}^n f(D_i|\lambda_{z_i}, E_i) \\ &= \prod_{j=1}^k \prod_{\{i: z_i=j\}} f(D_i|\lambda_j, E_i). \end{aligned} \quad (3)$$

At times, especially for the fixed k case described below, it is more convenient to work with (3) as it involves multiplications only, rather than additions and multiplications, as in (1).

The convenience of using the missing data formulation is that the posterior conditional distribution of the model parameters would be standard distributions. Also the augmented variables z allows us to see what component of the mixture the data points are assigned.

4.2. Gibbs Updates for Fixed k

We consider a mixture of Poissons where, conditional on there being k components in the mixture, we have

$$D_i \sim \sum_{j=1}^k w_j f(\cdot|\lambda_j, E_i).$$

The weights w_j sum to one, and are non-negative, so that

$$\sum_{j=1}^k w_j = 1, \text{ and } w_j \geq 0. \quad (4)$$

$$f(D_i|\lambda, z_i = j, E_i) \sim \text{Poisson}(\lambda_j E_i) \text{ with } P(z_i = j) = w_j$$

and

$$w \sim \mathcal{D}(\delta_1, \dots, \delta_k)$$

follows a Dirichlet distribution. We also make the additional assumption that the δ_j 's are equal to 1 so that $p(w)$ is a uniform distribution on the space described by (4). For the Poisson parameters λ_j , we take Gamma priors, so that

$$\lambda_j \sim \text{Gamma}(a, b) \quad j = 1, \dots, k$$

with the ordering constraint

$$\lambda_1 < \lambda_2 < \dots < \lambda_k, \quad (5)$$

to ensure that the components are identifiable. The ordering constraint is not necessary for the Monte Carlo algorithm to work. However it does avoid the problem of label switching, since otherwise, any permutation of the indices $\{1, \dots, k\}$ will result in the same posterior distribution.

Note that because we have the ordering constraint in Equation (5), the joint density of the collective λ is

$$p(\lambda|\alpha, \beta, k) = k! p(\lambda_1|\alpha, \beta) \cdots p(\lambda_k|\alpha, \beta) I_{\lambda_1 < \lambda_2 < \dots < \lambda_k}(\lambda).$$

When k is fixed and known, the factorial term $k!$ does not affect the MCMC algorithm since it can be absorbed into the normalising constant. However, in the variable k case, it must be noted, since it is a factor in the reversible jump acceptance probability. The joint density of all unknowns is

$$\pi(w, \lambda, z|D^n) \propto p(w|\delta) p(z|w) p(\lambda|\alpha, \beta) \mathbf{L}(D^n|\lambda, z, E^n). \quad (6)$$

With the missing data formulation, the likelihood term $\mathbf{L}(D^n|\lambda, z, E^n)$ can be written

$$\mathbf{L}(D^n|\lambda, z, E) = \prod_{i=1}^n \left(\frac{e^{-\lambda_{z_i} E_i} (\lambda_{z_i} E_i)^{D_i}}{D_i!} \right),$$

and

$$p(z|w) = \prod_{j=1}^k w_j^{n_j},$$

where $n_j = \#\{i|z_i = j\}$, is the number of observations allocated to component j . The prior distributions are

$$p(w|\delta) = \frac{\Gamma(\sum_{j=1}^k \delta_j)}{\prod_{j=1}^k \Gamma(\delta_j)} \prod_{j=1}^k w_j^{\delta_j - 1}$$

$$p(\lambda_i|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda_i^{\alpha-1} e^{-\beta \lambda_i}.$$

Using Bayes' theorem, we have the following posterior conditional distributions:

$$\pi(\lambda_j) \propto \lambda_j^{\alpha-1} e^{-\beta \lambda_j} \times \lambda_j^{\sum_{i|z_i=j} D_i} e^{-\lambda_j \sum_{i|z_i=j} E_i} I_{(\lambda_{j-1}, \lambda_{j+1})}(\lambda_j)$$

and

$$\pi(w|\delta, z) \propto p(w|\delta) p(z|w)$$

so that,

$$w \sim \mathcal{D}(\delta_1 + n_1, \dots, \delta_k + n_k),$$

where $n_j = \#\{i | z_i = j\}$. For z , we update the allocations using

$$P(z_i = j) \propto w_j f(D_i | \lambda_j, E_i) \quad i = 1, \dots, n; \quad j = 1, \dots, k,$$

so that,

$$p(z_i = j) = \frac{w_j f(D_i | \lambda_j, E_i)}{\sum_{j=1}^k w_j f(D_i | \lambda_j, E_i)}. \quad (7)$$

This follows from Equation (2). The Gibbs algorithm for fixed k is then (Robert and Casella, 1999)

Step 1: Simulate z_i from

$$p(z_i = j) \propto w_j f(D_i | \lambda_j, E_i) \text{ for } j = 1, \dots, k$$

and compute n_j , $n_j \bar{D}_j$, $n_j \bar{E}_j$ from

$$n_j = \sum_{i|z_i=j} (1) \quad n_j \bar{D}_j = \sum_{i|z_i=j} D_i \quad n_j \bar{E}_j = \sum_{i|z_i=j} E_i$$

Step 2: Simulate

$$\lambda_j \sim \text{Gamma}(\alpha + n_j \bar{D}_j, \beta + n_j \bar{E}_j) \mathbb{I}_{(\lambda_{j-1}, \lambda_{j+1})}(\lambda_j) \text{ for } j = 1, \dots, k$$

Step 3: Simulate

$$w \sim \mathcal{D}(\delta + n_1, \dots, \delta + n_k).$$

4.3. Reversible Jump MCMC

The Reversible jump algorithm is an extension of the Metropolis–Hastings algorithm. We assume there is a countable collection of candidate models, indexed by $M \in \mathcal{M} = \{M_1, M_2, \dots, M_k\}$. We further assume that for each model M_i , there exists an unknown parameter vector $\theta_i \in \mathbb{R}^{n_i}$ where n_i , the dimension of the parameter vector, can vary with i . Typically we are interested in finding which models have the greatest posterior probabilities and also estimates of the parameters. Thus the unknowns in this modelling scenario will include the model index M_i as well as the parameter vector θ_i . We assume that the models and corresponding parameter vectors have a joint density $\pi(M_i, \theta_i)$. The reversible jump algorithm constructs a reversible Markov chain on the state space $\mathcal{M} \times \bigcup_{M_i \in \mathcal{M}} \mathbb{R}^{n_i}$ which has π as its stationary distribution (Green, 1995). In many instances, and in particular for Bayesian problems this joint distribution is of the form

$$\pi(M_i, \theta_i) = \pi(M_i, \theta_i | \mathbf{X}) \propto \mathbf{L}(\mathbf{X} | M_i, \theta_i) p(M_i, \theta_i),$$

where the prior on (M_i, θ_i) is often of the form

$$p(M_i, \theta_i) = p(\theta_i | M_i) p(M_i)$$

with $p(M_i)$ being the density of some counting distribution.

Suppose now that we are at model M_i and a move to model M_j is proposed with probability r_{ij} . The corresponding move from $\boldsymbol{\theta}_i$ to $\boldsymbol{\theta}_j$ is achieved by using a deterministic transformation h_{ij} , such that

$$(\boldsymbol{\theta}_j, \mathbf{v}) = h_{ij}(\boldsymbol{\theta}_i, \mathbf{u}), \quad (8)$$

where \mathbf{u} and \mathbf{v} are random variables introduced to ensure dimension matching necessary for reversibility. To ensure dimension matching we must have

$$\dim(\boldsymbol{\theta}_j) + \dim(\mathbf{v}) = \dim(\boldsymbol{\theta}_i) + \dim(\mathbf{u}).$$

For discussions about possible choices for the function h_{ij} we refer the reader to Green (1995), and Brooks et al. (2003). Let

$$A(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j) = \frac{\pi(M_j, \boldsymbol{\theta}_j)}{\pi(M_i, \boldsymbol{\theta}_i)} \frac{q(\mathbf{v})}{q(\mathbf{u})} \frac{r_{ji}}{r_{ij}} \left| \frac{\partial h_{ij}(\boldsymbol{\theta}_i, \mathbf{u})}{\partial(\boldsymbol{\theta}_i, \mathbf{u})} \right| \quad (9)$$

then the acceptance probability for a proposed move from model $(M_i, \boldsymbol{\theta}_i)$ to model $(M_j, \boldsymbol{\theta}_j)$ is

$$\min \{1, A(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j)\}$$

where $q(\mathbf{u})$ and $q(\mathbf{v})$ are the respective proposal densities for \mathbf{u} and \mathbf{v} , and $|\partial h_{ij}(\boldsymbol{\theta}_i, \mathbf{u})/\partial(\boldsymbol{\theta}_i, \mathbf{u})|$ is the Jacobian of the transformation induced by h_{ij} . Green (1995) shows that the algorithm with acceptance probability given above simulates a Markov chain which is reversible and follows from the detailed balance equation

$$\pi(M_i, \boldsymbol{\theta}_i) q(\mathbf{u}) r_{ij} = \pi(M_j, \boldsymbol{\theta}_j) q(\mathbf{v}) r_{ji} \left| \frac{\partial h_{ij}(\boldsymbol{\theta}_i, \mathbf{u})}{\partial(\boldsymbol{\theta}_i, \mathbf{u})} \right|.$$

Detailed balance is necessary to ensure reversibility and is a sufficient condition for the existence of a unique stationary distribution. For the reverse move from model M_j to model M_i it is easy to see that the transformation used is $(\boldsymbol{\theta}_i, \mathbf{u}) = h_{ij}^{-1}(\boldsymbol{\theta}_j, \mathbf{v})$ and the acceptance probability for such a move is

$$\min \left\{ 1, \frac{\pi(M_i, \boldsymbol{\theta}_i)}{\pi(M_j, \boldsymbol{\theta}_j)} \frac{q(\mathbf{u})}{q(\mathbf{v})} \frac{r_{ji}}{r_{ij}} \left| \frac{\partial h_{ij}(\boldsymbol{\theta}_i, \mathbf{u})}{\partial(\boldsymbol{\theta}_i, \mathbf{u})} \right|^{-1} \right\} = \min \{1, A(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j)^{-1}\}.$$

For inference regarding which model has the greater posterior probability we can base our analysis on a realisation of the Markov chain constructed above. The marginal posterior probability of model M_i

$$\pi(M_i | \mathbf{X}) = \frac{p(M_i) f(\mathbf{X} | M_i)}{\sum_{M_j \in \mathcal{M}} p(M_j) f(\mathbf{X} | M_j)},$$

where

$$f(\mathbf{X} | M_i) = \int \mathbf{L}(\mathbf{X} | M_i, \boldsymbol{\theta}_i) p(\boldsymbol{\theta}_i | M_i) d\boldsymbol{\theta}_i$$

is the marginal density of the data after integrating over the unknown parameters $\boldsymbol{\theta}$. In practice we estimate $\pi(M_i | \mathbf{X})$ by counting the number of times the Markov chain visits model M_i in a single

long run after reaching stationarity. These between model moves described in this section are also augmented with within model Gibbs updates as given in Section 4.2 to update model parameters.

5. Reversible Jump Model Selection

To update the model order and thereby increase or decrease the number of components in the mixture, we use a combination of birth/death and split/merge moves as described below. We assume a uniform prior on the number of components k , so that

$$k \sim U\{1, \dots, k_{max}\},$$

where k_{max} is chosen to allow the algorithm to explore all feasible models. We set $k_{max} = 72$, the number of groups, as under our hypothesis, this is the maximum number of components in the mixture. $k = k_{max}$ only when the groups are all distinct. Setting $k_{max} = 72$ will allow for direct comparison of the empirical Bayesian and the mixture model approach.

Introducing a prior on the number of components k , we extend the joint density (6) of all parameters. Thus, now

$$\pi(k, w, z, \lambda | D^n) \propto p(k)p(w|\delta, k)p(\lambda|\alpha, \beta, k)p(z|k)\mathbf{L}(D^n|\lambda, z). \quad (10)$$

Note that the densities of the other model parameters now depend on k . In Sections 5.1 and 5.3, we describe in detail two algorithms which are used to simulate from this density. These algorithms are then combined with the fixed k updates of Section 4.2 to simulate from the density in Equation (10). Modelling mixtures with and without the Dirichlet process prior is considered by Green and Richardson (2001), who also considers the case of an unknown number of components. Alternatives to the reversible jump algorithm in this context do exist, see for example Dellaportas and Karlis (2001), who develop a semi-parametric sample based method to approximate a mixing density $g(\boldsymbol{\theta})$ based on the method of moments.

5.1. Split and Merge Moves

Note that the joint density in Equation (10) now depends on k . We use the split/merge method of Dellaportas et al. (1997) and Viallefont et al. (2002). Suppose we are at a configuration with k components, let

$$\boldsymbol{\theta}_k = \{(\lambda_1, w_1), \dots, (\lambda_k, w_k)\},$$

and suppose a move to increase the number of components is proposed. We select uniformly one of the current k components to be split. Suppose the j^{th} component, (λ_j, w_j) , is selected to be split into two components (λ_{j_1}, w_{j_1}) and (λ_{j_2}, w_{j_2}) such that $j_1 = j$ and $j_2 = j + 1$, the components originally numbered $j + 1, \dots, k$ are then renumbered $j + 2, \dots, k + 1$. The split is also designed so that the first two moments of the split component remains the same as the original component. Thus, we simulate u_1 and u_2 from densities defined on the interval $[0, 1]$. Usually, we use Beta

densities and set

$$\begin{aligned} w_{j_1} &= w_j u_1, \\ w_{j_2} &= w_j (1 - u_1), \\ \lambda_{j_1} &= \lambda_j u_2, \\ \lambda_{j_2} &= \lambda_j (1 - u_1 u_2) / (1 - u_1). \end{aligned}$$

Other choices for splitting and merging components are described in Viallefont et al. (2002). The proposed parameter is then

$$\boldsymbol{\theta}_{k+1} = \{(\lambda_1, w_1), \dots, (\lambda_{j-1}, w_{j-1}), (\lambda_{j_1}, w_{j_1}), (\lambda_{j_2}, w_{j_2}), (\lambda_{j+1}, w_{j+1}), \dots, (\lambda_k, w_k)\}.$$

If the ordering constraint in Equation (5) is not satisfied then the move is rejected immediately, as the reverse move in which we merge two adjacent components would not be possible. We can compute the Jacobian for this transformation as

$$\left| \frac{\partial \boldsymbol{\theta}_{k+1}}{\partial (\boldsymbol{\theta}_k, u_1, u_2)} \right| = \left| \frac{\partial (w_{j_1}, w_{j_2}, \lambda_{j_1}, \lambda_{j_2})}{\partial (w_j, \lambda_j, u_1, u_2)} \right| = \frac{\lambda_j w_j}{1 - u_1}. \quad (11)$$

For the reverse move, we select a pair of adjacent components j_1 and j_2 . Combining them, to get a new component labelled j , we set

$$w_j = w_{j_1} + w_{j_2}, \quad \lambda_j = \frac{w_{j_1} \lambda_{j_1} + w_{j_2} \lambda_{j_2}}{w_{j_1} + w_{j_2}},$$

by keeping the first two moments of the proposed and current configuration constant. We then sample a new set of allocation variables according to Equation (7). We also keep track of the probability of each allocation, so that $p_a(z)$ represents the probability of a given allocation. To compute $p_a(z)$, we first simulate z_i using Equation (7). For each i , the probability of that allocation is given by

$$p_a(z_i) = \frac{w_{z_i} f(D_i | \lambda_{z_i}, E_i)}{\sum_{j=1}^k w_j f(D_i | \lambda_j, E_i)}.$$

Finally, we compute the probability of all allocations by

$$p_a(z) = \prod_{i=1}^n p_a(z_i).$$

5.2. Acceptance Probability

The acceptance probability of a move of type $(k, \boldsymbol{\theta}_k) \Rightarrow (k', \boldsymbol{\theta}_{k'})$ is then $\min\{1, A_{k,k'}\}$, where

$$A_{k,k'} = \frac{\pi(k', \boldsymbol{\theta}_{k'})}{\pi(k, \boldsymbol{\theta}_k)} \times \frac{p(k' \Rightarrow k)}{p(k \Rightarrow k')} \times \frac{1}{q(u_1)q(u_2)} \times \left| \frac{\partial \boldsymbol{\theta}_{k'}}{\partial (\boldsymbol{\theta}_k, u_1, u_2)} \right|$$

$$A_{k,k'} = \frac{p(k')p(w'|\delta, k')p(\lambda'|\alpha, \beta, k')\mathbf{L}(D^n|\lambda', z')}{p(k)p(w|\delta, k)p(\lambda|\alpha, \beta, k)\mathbf{L}(D^n|\lambda, z)} \times \frac{p(z'|w', k+1)/p_a(z')}{p(z|w, k)/p_a(z)} \times \frac{p(k' \Rightarrow k)}{p(k \Rightarrow k')} \frac{1}{q(u_1)q(u_2)} \left| \frac{\partial \boldsymbol{\theta}_{k'}}{\partial (\boldsymbol{\theta}_k, u_1, u_2)} \right|.$$

With $k' = k + 1$ this becomes

$$A_{k,k+1} = \frac{p(k+1)}{p(k)} \times \frac{p(w'|\delta, k+1)}{p(w|\delta, k)} \times \frac{p(\lambda'|\alpha, \beta, k+1)}{p(\lambda|\alpha, \beta, k)} \times \frac{\mathbf{L}(D^n|\lambda', w')}{\mathbf{L}(D^n|\lambda, w)} \times \frac{p(k+1 \Rightarrow k)}{p(k \Rightarrow k+1)} \times \frac{1}{q(u_1)q(u_2)} \left| \frac{\partial \boldsymbol{\theta}_{k+1}}{\partial (\boldsymbol{\theta}_k, u_1, u_2)} \right|.$$

Now with a uniform prior on the number of components k and the weights w ,

$$A_{k,k+1} = \frac{\Gamma(k+1)}{\Gamma(k)} \times \frac{(k+1)p(\lambda_{j_1}|\alpha, \beta)p(\lambda_{j_2}|\alpha, \beta)}{p(\lambda_j|\alpha, \beta)} \frac{p(z'|w', k+1)/p_a(z')}{p(z|w, k)/p_a(z)} \times \frac{\mathbf{L}(D^n|\lambda', z')}{\mathbf{L}(D^n|\lambda, z)} \times \frac{m_{k+1}}{s_k} \times \frac{1}{q(u_1)q(u_2)} \times \left| \frac{\partial \boldsymbol{\theta}_{k+1}}{\partial (\boldsymbol{\theta}_k, u_1, u_2)} \right|,$$

where the ratio of Gamma terms comes from the ratio of the prior distributions on w' and w and

$$\frac{p(k+1 \Rightarrow k)}{p(k \Rightarrow k+1)} = \frac{m_{k+1}/(k+1-1)}{s_k/k} = \frac{m_{k+1}}{s_k}.$$

5.3. Birth and Death Moves

Suppose we are now at model M_k with k components, say

$$\boldsymbol{\theta}_k = \{(\lambda_1, w_1), \dots, (\lambda_k, w_k)\}. \quad (12)$$

If a move is proposed to increase the number of components by one, then we simulate

$$\tilde{w} \sim \text{Beta}(1, k) \text{ and } \tilde{\lambda} \sim \text{Gamma}(a, b),$$

independently. The proposed new component will then have weight \tilde{w} and the other weights are then scaled by a factor of $(1 - \tilde{w})$, so that the sum of the weights remain 1. The corresponding Poisson parameter for the proposed component is $\tilde{\lambda}$. Note that $\tilde{\lambda}$ is sampled from its prior distribution. The proposed component is then

$$\boldsymbol{\theta}_{k+1} = (\lambda_1, w_1/(1 - \tilde{w})), \dots, (\lambda_k, w_k/(1 - \tilde{w})), (\tilde{\lambda}, \tilde{w}). \quad (13)$$

Using this proposed value of $\boldsymbol{\theta}_{k+1}$, we also simulate proposed values for the allocations z' with model $k + 1$. Using the general form of the reversible jump acceptance probability, see for example Green (1995), the probability of changing the number of components to $k + 1$ is then $\min\{1, A_{k,k+1}\}$, where

$$A_{k,k+1} = \frac{\pi(k+1, \boldsymbol{\theta}_{k+1})}{\pi(k, \boldsymbol{\theta}_k)} \times \frac{p(k+1 \Rightarrow k)}{p(k \Rightarrow k+1)} \times \frac{1}{q(\tilde{w})q(\tilde{\lambda})} \times \left| \frac{\partial \boldsymbol{\theta}_{k+1}}{\partial (\boldsymbol{\theta}_k, \tilde{w}, \tilde{\lambda})} \right|.$$

Making the necessary substitutions yield

$$A_{k,k+1} = \frac{p(k+1)p(w'|\delta, k+1)p(\lambda'|\alpha, \beta, k+1)\mathbf{L}(D^n|\lambda', z')}{p(k)p(w|\delta, k)p(\lambda|\alpha, \beta, k)\mathbf{L}(D^n|\lambda, z)} \times \frac{p(z'|w', k+1)/p_a(z')}{p(z|w, k+1)/p_a(z)} \times \frac{p(k+1 \Rightarrow k)}{p(k \Rightarrow k+1)} \times \frac{1}{q(\tilde{w})q(\tilde{\lambda})} \times \left| \frac{\partial \boldsymbol{\theta}_{k+1}}{\partial(\boldsymbol{\theta}_k, \tilde{w}, \tilde{\lambda})} \right|. \quad (14)$$

Using Equations (12) and (13) we then have the Jacobian

$$\frac{\partial \boldsymbol{\theta}_{k+1}}{\partial(\boldsymbol{\theta}_k, \tilde{w}, \tilde{\lambda})} = (1 - \tilde{w})^{k-1}.$$

If we denote the probability of a birth when there are k components by b_k , and the probability of a death by d_k , with $b_k + d_k = 1$, then

$$\frac{p(k+1 \Rightarrow k)}{p(k \Rightarrow k+1)} = \frac{d_{k+1}/(k+1)}{b_k},$$

since for the move to be reversible we would then be able to kill $k+1$ components in the new model, each with equal probability. Substituting these values in Equation (14), the ratio $A_{k,k+1}$ reduces to

$$A_{k,k+1} = \frac{\Gamma(k+1)}{\Gamma(k)} \times (k+1)p(\tilde{\lambda}) \times \frac{\mathbf{L}(D^n|\lambda', z')}{\mathbf{L}(D^n|\lambda, z)} \times \frac{p(z'|w', k+1)/p_a(z')}{p(z|w, k)/p_a(z)} \times \frac{d_{k+1}/(k+1)}{b_k} \frac{1}{q(\tilde{w})q(\tilde{\lambda})} \times (1 - \tilde{w})^{k-1},$$

which on substituting $q(\tilde{\lambda}) = p(\tilde{\lambda})$ and $q(\tilde{w}) = k(1 - \tilde{w})^{k-1}$ further reduces to

$$A_{k,k+1} = \frac{p(z'|w', k+1)/p_a(z')}{p(z|w, k)/p_a(z)} \frac{\mathbf{L}(D^n|\lambda', z')}{\mathbf{L}(D^n|\lambda, z)} \times \frac{d_{k+1}}{b_k}. \quad (15)$$

For a proposed death move, the acceptance probability is then

$$\min\{1, A_{k,k+1}^{-1}\}.$$

Even though the algorithm simulates new values of for the allocations when proposing to move, it is not necessary to carry the allocations along. For between-model moves, we could replace the missing data formulation by noting that

$$\frac{p(z|w, k)}{p_a(z)} \mathbf{L}(D^n|\lambda, z) = \mathbf{L}(D^n|\lambda, w).$$

Thus we could update model parameters using a scheme which does not require conjugacy; see for example Cappé et al. (2003a,b).

6. Results

We now present some numerical results for this dataset based on the model described in Section 4 and using the algorithms described in Section 5.

Table 1. Posterior Model Order.

Model Order k	Posterior Probability $\pi(k D^n, E^n)$
1	0.00000
2	0.59485
3	0.29058
4	0.08588
5	0.02258
6	0.00448
7	0.00104
8	0.00034
9	0.00026
10	0.00000

Table 2. Acceptance Rates.

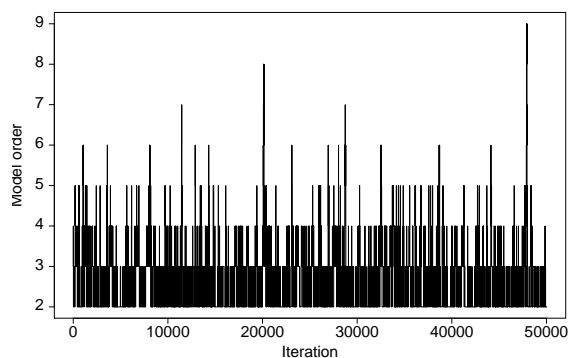
Scheme	Acceptance Rate
Birth/Death	0.077
Split/Merge	0.055
Birth/Death and Split/Merge	0.066

Table 1 shows the posterior model probabilities calculated from the reversible jump algorithm by counting the proportion of times the algorithm visits each model. A plot of the number of components as the chain evolves is shown in Figure 3(a). The results show clearly that the number of components has a posterior mode at $k = 2$. Also, the model with $k = 1$ component is never visited. If the algorithm is started with $k = 1$ then immediately it jumps to $k = 2$ and never returns to $k = 1$. Since more than 88% of the posterior probability mass is placed on the models with 2 or 3 components, we discuss those models in detail in Section 6.2. The between-model acceptance rates were 7.7% and 5.5% for the birth/death and split/merge moves, respectively. The total acceptance rate when there is equal probability of proposing a birth/death move or a split/merge move, is 6.6%. These results are tabulated in Table 2.

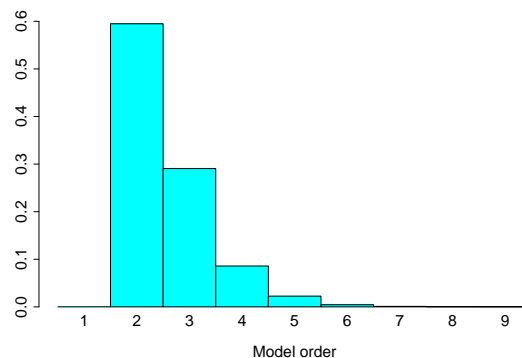
To assess convergence of the algorithm, we simulated 4 chains using different starting values and different random number seeds for a total of 100000 iterations. Both the χ -square and Kolmogorov-Smirnov diagnostics are computed. These diagnostics are plotted in Figure 4.

6.1. Comparing the Model Move Schemes

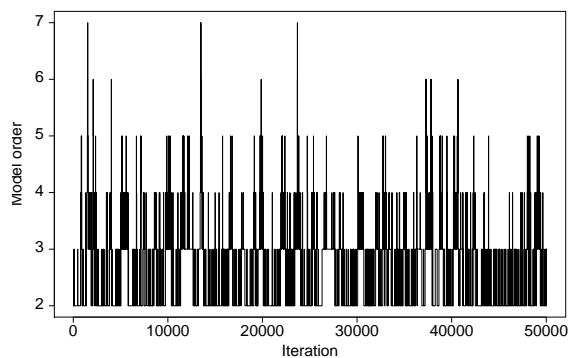
A comparison of the individual acceptance probabilities shows that the between model moves are accepted with a larger rate for the birth death scheme compared with the split merge scheme. This might not always be true, as other split merge schemes may be proposed (Viallefont et al., 2002). It is interesting to note that although the birth and death rates are higher than the split and merge rates, the combined scheme seems to mix better than either scheme implemented alone. Even though the birth and death scheme have a higher acceptance rate for between-model moves, the excursions away from the values of highest posterior density, $k = 2$ and $k = 3$, are longer than for the combined scheme or the split and merge scheme since. This is because when proposing parameters independently from the prior, areas of low probability can be proposed, whereas, with the split and merge scheme, areas of low probability mass will generally be rejected. Based on the results presented here, the birth/death method would be the preferred algorithm.



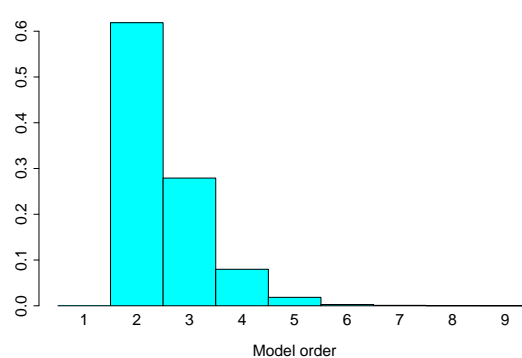
(a) Birth/death and split/merge model trace.



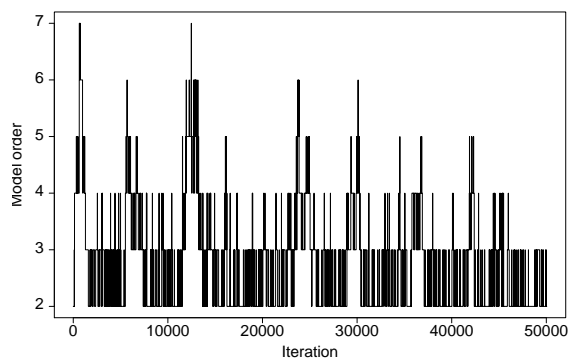
(b) Birth/death and split/merge histogram.



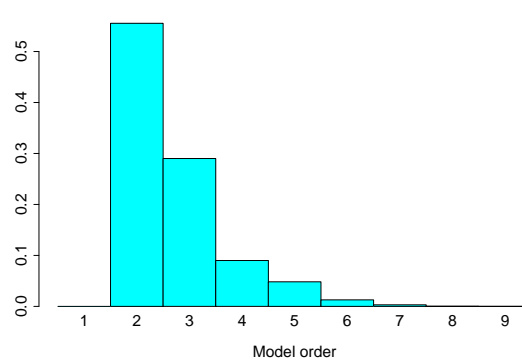
(c) Birth/death model trace.



(d) Birth/death histogram.



(e) Split/merge model trace.



(f) Split/merge histogram.

Fig. 3. Left: Model trace indicator. Right: Histogram of posterior model order.

Table 3. Parameters Estimates Conditional on $k = 2$.

	Estimate	95% HPD Interval
λ_1	0.731	(0.626, 0.839)
w_1	0.636	(0.428, 0.821)
λ_2	1.896	(1.557, 2.249)
w_2	0.363	(0.178, 0.571)

Table 4. Parameters Estimates Conditional on $k = 3$.

	Estimate	95% HPD Interval
λ_1	0.462	(0, 0.770)
w_1	0.299	(0.000, 0.656)
λ_2	1.115	(0.625, 1.692)
w_2	0.495	(0.157, 0.797)
λ_3	2.481	(1.570, 3.366)
w_3	0.204	(0.002, 0.464)

6.2. Detailed Results for $k = 2$ and $k = 3$

A histogram plot of the model indicator from the reversible jump algorithm of Section 5 shows that the most plausible model generating the claims in the portfolio is a mixture of two Poisson distributions. In this section we look further at the results, conditional on their being only two components, or three components, in the mixture.

Recall the missing data formulation introduced in Section 4 for the number of components conditional on $k = 2$, we observed the posterior distribution of z at each iteration when $k = 2$. A study of values of z will tell us how the data has been allocated to the components and therefore, which data points have been generated from either the first Poisson distribution or the second Poisson distribution. This information, along with further information from the portfolio, will help insurance companies classify groups of life insurance portfolios. The parameter estimates are shown in Table 3.

Similar results for the posterior parameter estimates, conditional on there being three components in the mixture, are given in Table 4.

Figure 5 show the posterior probability of each data point being allocated to a particular component of the mixture, conditional on $k = 2$ and conditional $k = 3$, respectively.

7. Summary

We present a model for heterogeneity in group life insurance. We show that the assumption of identical heterogeneity for all groups under consideration, may not necessarily hold. In this case, it is necessary to put similar groups together for further analysis. We employ a non-parametric approach and apply reversible jump methods to determine the number of components in the mixture. An extension of the current work to the case where claims are grouped, such as (Walhin and Paris, 1999, 2000), would therefore be appropriate.

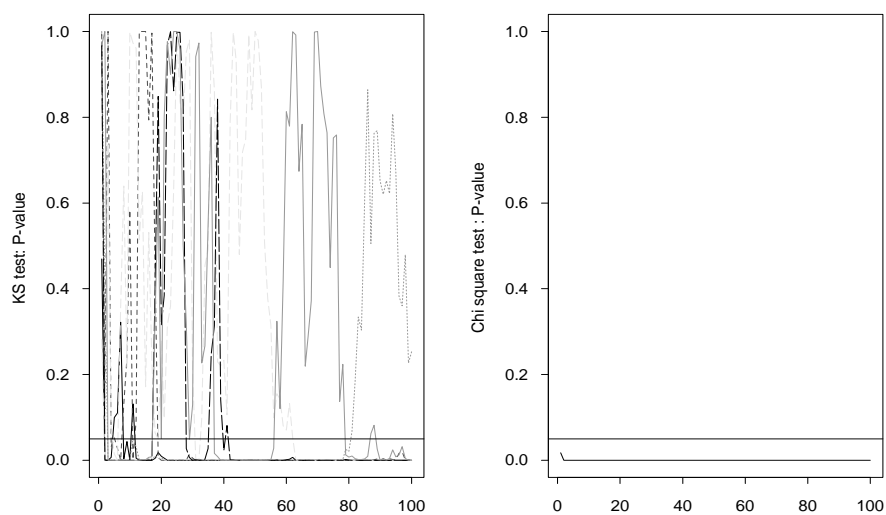


Fig. 4. Convergence diagnostics.

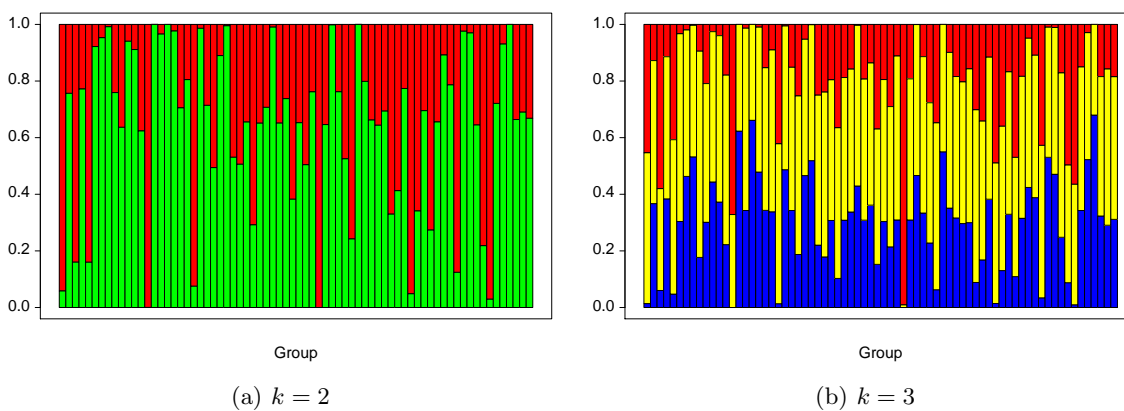


Fig. 5. Probability (vertical axis) of data from Group i (horizontal axis) being assigned to individual components conditional on $k = 2$ (left) and $k = 3$ (right).

References

- Brooks, S. P., P. Giudici, and G. O. Roberts (2003). Efficient construction of reversible jump MCMC proposal distributions (with discussion). *Journal of the Royal Statistical Society, Series B* **65**(1), 3–55.
- Cappé, O., C. P. Robert, and T. Rydén (2003a). CT/RJ-Mix: Transdimensional MCMC for Gaussian mixtures (available as C source code).
- Cappé, O., C. P. Robert, and T. Rydén (2003b). Reversible jump, birth-and-death and more general continuous time Markov chain Monte Carlo samplers. *Journal of the Royal Statistical Society, Series B* **65**(3), 679–700.
- Carlin, B. P. and S. Chib (1995). Bayesian Model Choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B* **57**, 473–484.
- Casella, G., C. P. Robert, and M. T. Wells (2000). Mixture Models, Latent Variables and Partitioned Importance Sampling. Technical Report 2000–03, CREST, INSEE, Paris.
- Dellaportas, P. and D. Karlis (2001). A simulation approach to nonparametric empirical Bayes analysis. *International Statistical Review* **69**(1), 63–79.
- Dellaportas, P., D. Karlis, and E. Xekalaki (1997). Bayesian analysis of finite Poisson mixtures. Technical report, Athens University of Economics and Business.
- Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**(4), 711–732.
- Green, P. J. and S. Richardson (2001). Modelling heterogeneity with and without the Dirichlet process. *Scandinavian Journal of Statistics* **28**(2), 355–375.
- Green, P. J. and S. Richardson (2002). Hidden Markov Models and Disease Mapping. *Journal of the American Statistical Society* **97**(460), 1055–1070.
- Haastrup, S. (2000). Comparison of Some Bayesian Analyses of Heterogeneity in Group Life Insurance. *Scandinavian Actuarial Journal* **2000**, 2–16.
- McLachlan, G. and D. Peel (2000). *Finite Mixture Models*. Wiley-Interscience.
- Norberg, R. (1989). Experience Rating in Group Life Insurance. *Scandinavian Actuarial Journal* **1989**, 194–224.
- Phillips, D. B. and A. F. M. Smith (1996). Bayesian model comparison via jump diffusions. In W. R. Gilks, S. Richardson, and D. J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice*, pp. 215–239. Chapman and Hall.
- Robert, C. P. and G. Casella (1999). *Monte Carlo Statistical Methods*. Springer.
- Stephens, M. (2000). Bayesian analysis of mixture models with an unknown number of components—an alternative to reversible jump methods. *Annals of Statistics* **28**(1), 40–74.

- Titterton, D. M., A. F. M. Smith, and U. E. Makov (1990). *Statistical Analysis of Finite Mixture Distributions*. New York: Wiley.
- Tremblay, L. (1992). Using the Poisson inverse Gaussian in bonus–malus systems. *ASTIN Bulletin* **22**(1), 97–106.
- Viallefont, V., S. Richardson, and P. J. Green (2002). Bayesian analysis of Poisson mixtures. *Journal of Nonparametric Statistics* **14**(1–2), 181–202.
- Walhin, J. F. and J. Paris (1999). Using mixed Poisson processes in connection with bonus–malus systems. *ASTIN Bulletin* **29**(1), 81–99.
- Walhin, J. F. and J. Paris (2000). The true claim amount and frequency distributions within a bonus–malus system. *ASTIN Bulletin* **30**(2), 391–403.